DATA MINING THROUGH RELATIONALDATABASE

[1]Ayush Kumar, [2]Neelesh Jain, [3]Neeraj Gupta

[1]Student, [2]Professor, [3]Professor

[1,2,3]Department of Computer Science & Engineering

[1,2,3]SAM College of Engineering and Technology, Bhopal

Abstract:- With the wide availability of huge amounts of data and the imminent demands to transform the raw data into useful information and knowledge, data mining has become an important research field both in the database area and the machine learning areas. Data and Information or Knowledge has a significant role on human activities. Data mining is the knowledge discovery process by analyzing the large volumes of data from various perspectives and summarizing it into useful information. Due to the importance of extracting knowledge/information from the large data repositories, data mining has become an essential component in various fields of human life. Advancements in Statistics, Machine Learning, Artificial Intelligence, Pattern Recognition and Computation capabilities have evolved the present day?s data mining applications and these applications have enriched the various fields of human life including business, education, medical, scientific etc. Hence, this paper discusses the various improvements in the field of data mining from past to the present and explores the future trends. Data mining is defined as the process to solve problems by analyzing data already present in the database and discovering knowledge in the data. Database systems provide efficient data storage, fast access structures and a wide variety of indexing methods to speed up data retrieval. Machine learning provides theory support for most of the popular data mining algorithms. Database combines properties of these two areas to improve the scalability of Weka, which is an open source machine learning software package. Weka implements most of the machine learning algorithms using main memory based data structure, so it cannot handle large datasets that cannot fit into main memory. Database is implemented to store the data into and access the data from DB2, so it achieves better scalability than Weka. However, the speed of Database is much slower than Weka Database because secondary storage access is more expensive than main memory access. In this thesis we extend Database with a buffer management component to improve the performance of Database . Furthermore, we increase the scalability of Database even further by putting further data structures into the database, which uses a buffer to access the data in database. Furthermore, we explore another method to improve the speed of the algorithms, which takes advantage of the data access properties of machine learning algorithms.

Keywords: Applications of Data Mining, Business Intelligence, Data Mining, Data Presentation, Database Systems

I. INTRODUCTION

Since the inception of information storage, the ability to sift through and analyze huge amounts of information was a dream sought out for in many ways and through different ways. With the advent of electronic and magnetic data storage, rational databases emerged as one of the efficient and widelyused method to store data. Data stored in such large databases are not always comprehendible by humans, it needed to be filtered and analyzed first. Stored records are raw amounts of data poor in information, not only is it large and seamlessly irrelevant but also continuously increasing, updating and changing. Here is where data mining and Presentation comes into the picture. Data mining and Presentation s are knowledge discovery tools used for autonomous analysis of data stored in large sets in many different ways. Large data sets ofdata cannot possibly be analyzed manually; mining tools and Presentation provide automated means to comprehend such data sets. Data mining is defined as the automated process of finding patterns, relationships, and trends in the data set. On the other hand, data Presentation is the process of visually representing the data set in a meaningful and comprehendible manner. In fig. 1, the figure shows what Data Mining is and is not.

Data mining is a knowledge discovery process; it is the analysis step of knowledge discovery in databases or KDD for short. As an interdisciplinary field of computer science, it involves techniques from fields such as artificial intelligence AI, machine learning, probability and statistics theory, and business intelligence. As in actual mining, where useful substance is mined out of large deposits hidden deep with mine. Data mining mines meaningful and hidden patterns, and it’s highlyrelated to mathematical statistics. Though utilizing pattern recognition techniques, AI techniques, and even socio-economic aspects are taken into consideration. Data mining is used in today’s ever- growing databases to

achieve business superiority, finding genome sequences, automated decision making, monitoring and diagnosing engineering processes, and for drug. Discovery and diagnosis in medical and health care. Data Mining, as with other Business Intelligence tools, efficiency is affected bythe Data Warehousing solution used.

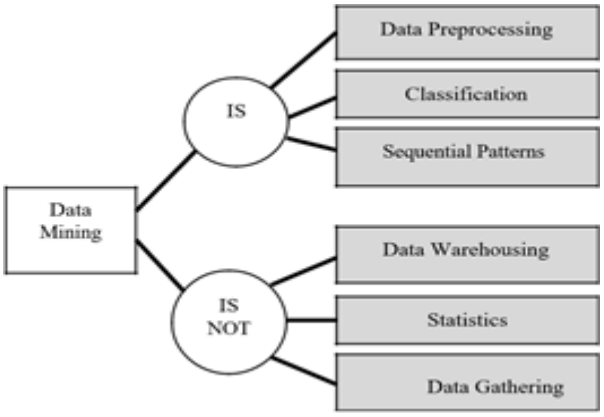

Figure 1: What is Data Mining?

Data Presentation is a data mining application considered as an information-modeling paradigm, in which seamlessly random data may be represented in an appealing graphical manner.

Presentation of collected data can be found as early as the middle of the 19th century, where Dr. John Snow made a map of central London, pinpointing the locations of the possible sources of the cholera and its victims. Thus allowed for the detection of thehidden relation of the alleged sources of cholera (the water pumps) and its victims, and helped in ridding of the disease, other examples are also given in. Presentation can be divided in to seven main subfield according to Frits, Presentation algorithms, volume Presentation , information Presentation , multi-resolution methods, modeling techniques, and architecture and interaction techniques. Human beings understand and comprehend graphics more easily than numbers and letters. Human brains can interpret graphs, charts, icons and models quicker than numbers in tables; this is in contrast to computers, where numerical representation is perceived more efficiently. For example, a pie chart showing the classification of a university student will be understood quicker than the same data represented in a table, as Fig. 2 shows. Presentation of such data helps the human brain figure out and perceives such knowledge hidden in the data. The goal of data Presentation is to not only summarize the large dataset, but also provide a better way of exploring the knowledge hidden and waiting to be found there automatically and autonomously. Presentation of datasets helps in explicitly showing proximity, enclosure, similarity, connection, and continuity. Analysis of data through Presentation is further divided into two main categories, the Exploratory Data Analysis EDA, and Qualitative Data Analysis QDA;

II. BACKGROUND

The notion of automated discovery tool in a large data set has prevailed in the development of data storage technologies. Tracing the roots of data mining to the early days of mathematical regression and probability theories in the eighteenth century, we can see that mathematical models such as regression and Bayesian theories provided means of analyzing large data sets effectively. With electronic computers taking the exclusive position for data storage in the twentieth century, early commercial computers quickly over took manual and other means of data storage. By the 1950s, early high level languages were developed; this development dramatically changed how humans interact with computers. Computerized data storagewas not only used for storage but also for querying. Further on after the advancements in both hardwareand software, rational database systems RDBS were developed. Structured Query Languages SQLs were used for semi-automatic acquisition of knowledge through querying the data storage, although tedious programming and substantial efforts have to be done.

III. RELATED WORK

In the author reviews several interactive Presentation techniques that are used in the context of data mining. The paper also retrospectively defines Presentation techniques in the world of data mining; these can be defined as expressing data sets to discover trends, for *exploration*, or can visualize the workings of complex data mining processes, for *comprehension*. The paper focuses on data Presentation, while in our survey we shall review both data mining and data Presentation and their integration as one field.

Authors C. Romero and S. Ventura of give a survey data mining techniques in the field of education. Not just in e-learning but also in traditional class rooms. Data mining can help in improving educational courses through knowledge discovery of facts in the past history of a specific course. These include: feedback for the educators such as effectiveness of content, students' classifications, and mistakes in the teaching process, feedback for the students such as suggesting helpful educational content available for them. The paper surveys data mining and a few data Presentation techniques used in education such as classification, text mining, sequential patterns and Presentation. In our survey, data mining and Presentation techniques, trends,

and application will be discussed not only for education but for a wider range of fields.

IV. TASKS AND TECHNIQUES

Data mining and data Presentation were developed from mathematical methods of pattern recognition and probabilistic theories to deal with unstructured, time varying, and fuzzy data in huge amounts. Suchtechniques allowed for finding correlations,relations and assertions. We shall touch upon some of the main tasks associated with data mining and Presentation and the techniques to achieve such tasks that are popular in the field of data mining andPresentation in the following paragraphs.

Clustering is the technique of grouping of several objects unto groups of similar attributes in order to simplify large, complex sets. Clustering is a learning technique and therefore it has no correct answer. Clustering can be hierarchical and non-hierarchical. Hierarchical clustering clusters groups of data in size (can be from small to large or vice versa), and it comes in two flavors, Agglomerative and Divisive. The first clusters each record alone, and then merges clusters together. The second, does the opposite, it starts with one full cluster and then subdivides the cluster. The non-hierarchical clustering has two flavors as well; the difference here is that no hierarchic clusters are used. The first type is the single pass methods, where the database is scanned once to create the cluster. The second type is the relocation method, where records are relocated from one cluster to another for optimization. Several passes against the database may be used, as opposed to the single pass methods.

Sequential Patterns the use of sequential pattern algorithms on sets of sequential data (e.g. bills madeon the same month). The goal is to find a trend or pattern that happen in sequence. Rule induction task is used to find hidden if-then rules in the dataset. These rules are based on statics analysis and probabilistic models. Derived if-then rules are further used in analyzing the dataset in the future.

V. PRESENTATION OF LARGE DATABASES

As discussed previously in this paper, data Presentation is an important application that helps to convey knowledge mined *graphically*. As human beings, we are more familiar with drawings, icons, and graphs then we are with numbers and tables. Raw numerical data or even alphanumerical data can be represented in a map; chart, bars, pie chart, or even a histogram to visually identified and convey important trends and correlations visually. Data mining in large databases is still a difficult task; due to the fact of the huge amount of raw dataneed to be processed. A turnaround is to partition the data into sets, and tackle them individually. This makes supporting tools such as *Presentation Tools*needed. Data Presentation is considered with two kind of analysis, first, Exploratory Data analysis EDA and model Presentation [30], second is the Qualitative Data Analysis QDA.

By EDA, it is meant the careful exploration of the data set graphically to identify a pattern, a recurring trend or behavior thaat connects different views or Presentation. EDA heelps to identify patterns without preconceived knowledge, hypothesis, or suggested models used o n the data set. Model Presentation is the use of predefined models, such as XY charts, 3D plots, or box plots to model the data. Usually Presentation of data plays on the key idea that human beings are more capable in analyzing and understanding graphs than digits andletters. Figure 2 include a very simple, yet effective example of tabular versus visualized data sets. Presentation s such as Venn diagrams and clustering help the observers see grouping and partitions in a dataset more easily than rows of alphanumerical records. QDA on the other hand, is the analysis of non-numerical data. QDA is considered with database containing images, text, links, or other kinds of data that is not numerical or alphanumerical.

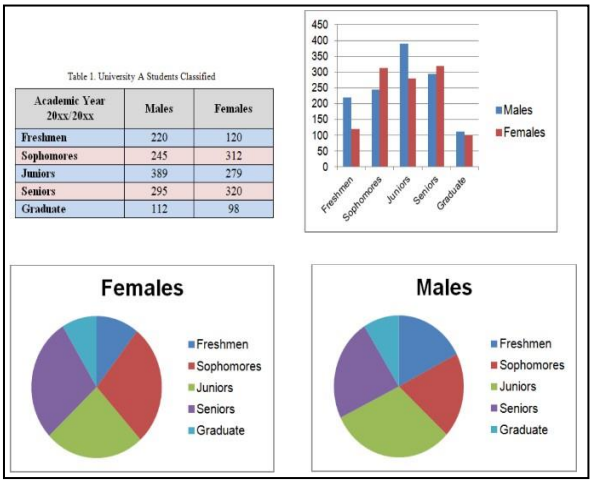

Figure 2 (a): Histogram RepresentationFigure 2 (b): Pie-Chart Representation

Figure 2: Different Data Representation

Prominent Presentation techniques used to visualize large datasets is charting. Charts or namely pie charts are the most common form of data Presentation. Pie charts ar e both easy to understand and an elegant and fast delivery method. As most people are familiar with pie charts, they convey information relatively in a fast and direct way, see Fig2. Large database consists of millions of data records that are updated frequently, an example of such databases are Geographical Information Systems GISs. Presentation is used in GISs to visualize dataset.

VI. APPLICATION OF DATA MINING

Applications of data mining vary, depending on the nature of the data to be mined. Since its inception data

mining was used in various other fields. The classical application of data mining encompasses statistical and probabilistic applications. These classical applications included for example, population census studies, biosphere analysis, and marine life and oceanography. As a prominent use for data mining is data Presentation , we have touched upon this application in the previous section; we focused on Presentation since its importance as an application for data mining. We have selected other important applications of data mining used widely today. Many of these applications have branched out to become separate but related fields.

6.1 Spatial Data Mining

Spatial databases are databases that have unique data; this data is about space and geometry, such as the coordinates of earth, maps, and satellite data. This data is in the form of geological or geographical data. Such databases are extremely large and the data seems for the most part unrelated, and without any signs or correlations. Data mining is a natural candidate to find logic and make sense of such data. Data Presentation, which was discussed earlier, is another tool of data mining heavily used in spatial databases.

6.2 Business Intelligence

Data mining help business intelligence in manyways and for that, it is one of the fundamental tools of business intelligence. Business intelligence's (BI) goals are to gain a competitive advantage over competitors, increase productivity and effectiveness of current business operations, and to maintain a balance and control of risk management. Business intelligence is a usual task of any Enterprise Resource Planning ERP solution. Business-intelligence mine habits and trends of customers' data stored as records through internet cookies and sales profiles. This mining helps in discovering the customers' segmentations, and demographics. Data mining provides market basket analysis; items purchased together are identified and in turn bundled and advertised together. Anomalies can be also caught using intelligent mining tools; such tools mine the transactions and try to extract anomalies. Anomalies may be deliberate, such as fraudulent transactions or they could be unintentional, a glitch or bug in the program or just an odd transaction that may never be presented again in the entire database. Fraudulent transactions are caught due to their recurring characteristics,such as credit card theft, identity thefts or account hackings.

6.3 Text Mining

Another widely used application of data mining is text mining. Text mining deals with textual data rather than records stored in a regular database. It is defined as an automated discovery of hidden patterns, knowledge, or unknown information from textual data. Most of data found on the World Wide Web WWW is text, after distilling the multimedia elements; most of knowledge out there is text. Text mining utilizes different techniques and methodologies, mostly linguistic and grammatical techniques, such as the Natural Language Processing NLP. Techniques of text mining originated from computational linguistics, statistics, and machine learning, such techniques were developed to make machines, specifically speaking computers, understand human language.

6.4 Web Mining

With the revolution of the Internet that have changed how databases are used, this revolution brought the term of web mining. Web mining is considered as a subfield of data mining, it's regarded as the vital web technology that is used heavily to monitor and regulate web traffic. Web mining is further divided into three main sub groups, web content mining, web structure mining, and web usage mining. Web content mining is the mining of content found on the web, this include metadata, multimedia, hyperlinks and text. Web structure mining is considered with the semantics and hyperlinks making up a website or a network. Web structure mining are usually is used by search engines to 'crawl' the web and find all possible links forming a network. Web usage mining is considered with the traffic patterns in the World Wide Web WWW. Most of the data is mined from the web servers and web proxies. Web servers log most traffic, such logs are the data needed to construct an overview map of the traffic coming and going to that web site.
Web mining is used in Information Retrieval IR systems, such as search engines. Web mining is also used in web trafficking measures, were traffic is traced and monitored. But for the most times, web mining is used for business intelligence [47], as it can search the web with all its fuzziness to retrieve business oriented information from the web.

VII. TOOLS

Data mining tools are basically software packages, whether integrated packages or individual packages. These sophisticated software tools often require special data analysts. Such analysts are trained to use such tools , as data mining itself is not a straightforward process. It is worth mentioning that data mining tools need a substantial investment in hardware and software, as well as human resources. Deployment of data mining tools and packages is also an overwhelming task, in size and management, as it needs careful planning and management. In the next paragraphs, we shall look into some of the used data mining tools and data Presentation tools.

7.1 Data Mining Tools

Data mining tools are also called *siftware*, for the sole reason that they 'sift' through the dataset. Data mining tools varies depending on level of theirsophistication and projected level of accuracy. In 2008, the global market for business intelligence software, data mining centric software, reached over 7.8 billion USD, a vast amount. IBM *SPSS* is an example of business intelligence software package; it is integrated data mining software with diverse business intelligence capabilities. IBM also provides online services for web mining, these services are called *Surfaid* Analytics; they provide sophisticated tools for web mining. Other data mining with business intelligence capabilities is *Oracle Data Mining*, a part of the company's flagship RDBMS software suite. SAS also offers its *SAS Enterprise Miner*, as a part of its enterprise solutions. SAP, a world-renowned business solution provider, offers world known ERP solutions along with providing other mining tools software that can be integrated into their ERP solutions. Other software companies include Microsoft; it offers *SQL Server Analysis Services*, a platform dependent solution integrated in Microsoft SQL platform for Microsoft Windows Server. Microsoft also offers a less sophisticated product, namely the *PowerPivot*, a mining tool for small and middle size enterprises, with limitations and ease of use to match with its nature of use. Open source mining tools exist; they include the Waikato Environment for Knowledge Analysis or WEKA.

7.2 Data Presentation tools

For Data Presentation tools, we have checked IBM's *Parallel Visual Explorer*. This software package is used for market analysis, oil exploration, engineering and aerospace applications, and agriculture to name a few.

For medical fields, Parallel Visual Explorer is used to analyze various effects of treatments on the immune system. It helps in visualizing many different diverse effects on the patients' immune system. For manufacturing, this tool helps in monitoring the processing parameters. Process parameters are vital for effective streamlined production. For agricultural usages, this tool helps in determining which seed to plant by analyzing thesoil parameters with taking in consideration the weather conditions. Finally, Parallel Visual Explorer is also used for market research such as providing visual aids to help market analyst find customers trends, habits, and buying sprees.

An interesting Presentation tool is Cave5D. It's a data Presentation tool developed by the university of Wisconsin-Madison. The inventors of this tool are Glen Wheless, Cathy Lascara, from the center for Pacific Oceanography, with Bill Hibbard and Brian Paul back in 1994. This software ran as a package for the Vis5D software. Cave5 provides interactive 3D, time variable Presentation of dataset in a virtual environment. Cave5D integrates Vis5D's libraries and framework; it uses its graphical libraries to model the dataset.

VIII. METHODOLOGY

Data Collection and Selection: Gather relevant datasets from various sources, and select anappropriate relational database management system (RDBMS) for efficient data storage and management. Ensure data quality and relevance foraccurate analysis.

Data Preprocessing: Apply data cleaning techniques to handle missing values, outliers, and duplicate entries. Use data transformation methods, such as normalization and encoding, to standardize attributes. Implement feature selection techniques like Recursive Feature Elimination or Principal Component Analysis to identify the most significantattributes and reduce dimensionality.

Association Rule Mining: Utilize the Apriori or FP-Growth algorithms to uncover relationships and patterns between attributes in the relational database, helping to identify hidden correlations anddependencies.

Classification: Implement supervised learning algorithms like Decision Trees, Naïve Bayes, and Support Vector Machines to classify instances in the database. These techniques enable predictive modeling and trend analysis, facilitating informed decision-making.

Clustering: Apply unsupervised learning algorithms such as K-Means, DBSCAN, or Hierarchical Clustering to group similar instances. Clustering helps reveal hidden patterns, detect anomalies, and segment data for further analysis ortargeted actions.

Evaluation and Validation: Assess the performance and accuracy of the data mining algorithms using cross-validation, confusion matrices, and performance metrics such as precision, recall, and F1-score. Proper evaluation ensures the reliability and generalizability of the mining results.

Visualization and Interpretation: Develop interactive visualization tools and methodologies to represent the mined data and patterns, allowing users to explore and gain insights from the relationaldatabase. Techniques like parallel coordinates, scatter plot matrices, and tramps can aid inunderstanding complex data relationships.

Performance Optimization: Investigate efficient query processing and indexing techniques for relational databases to reduce the computationaloverhead of data mining algorithms. By optimizing database design and utilizing parallel processing, the overall performance and scalability of data mining processes can be improved.

Privacy-preserving Data Mining: Develop secure and privacy-preserving algorithms for data mining in relational databases, addressing concerns related to data confidentiality and compliance with data protection regulations. Techniques like k- anonymity, l-diversity,

and differential privacy can be employed to ensure the privacy of sensitive information while still enabling valuable data mining.

Real-world Application: Apply the proposed methodology to a specific domain, such as healthcare, finance, or e-commerce, to demonstrate its practical utility and effectiveness. By analyzing real-world data, the strengths and limitations of the methodology can be identified, allowing for further refinement and improvement.

IX. FUTURE TRENDS

Future trends for data mining lie in the hands of innovation and scientific breakthrough. As data mining is both a difficult problem, and a relatively new problem that incorporates many interdisciplinary fields. We shall see some new trends that will shape the way that data mining will be used in the upcoming future. Presentation tools are also witnessing a rise, credited to the newer technologies in human-computer interactions.

9.1 Cloud Computing Based Data Mining

A relatively new trend in utilizing and benefiting from data mining tools for middle-sized and small enterprises, incapable of supporting a full-fledged data mining solution, is *cloud computing* based data mining tools. Because small and middle-sized enterprises usually lack the infrastructure and budget available for large enterprises, they tend to try this new cost effective trend. Cloud computing promises to provide data mining tools benefits at relatively lower costs form such small or middle sized enterprises. Cloud computing provides web data warehousing facilities, were the actual data warehouse application is outsourced and accessed entirely through the World Wide Web. Cloud based data mining also provides sophisticated mining analysis of the dataset, comparable to actual data mining software, as the enterprise specifies and demands.

Aside from lowering the costs of the data mining software tools infrastructure, cloud based mining also provides expertise that is not available in such middle-sized and small enterprises. Most cloud based data mining providers tend to have data experts, data analysis, and a broader experience with data mining then their clientele. Usually start- ups or entrepreneur level enterprises lack not only the financial resources but also the human resources and expertise in the Information Technology IT field, not to mention in the data analysis field.

9.2 Data Conditioning Tools

Data conditioning is currently a technique that is not only meant for data mining. It is used for intelligent routing, privacy and protection as well as for data mining. As data grows today in unprecedented rate, the need to clean up the huge piles of data is necessary. Reports suggest that more than 80% of enterprises data are unstructured and fuzzy data. The other goal of data conditioning is to elevate or at least minimize the interference of IT people. This would quicken the BI step, and in turn make it ubiquitous for the end-users, whether business or science users.

The key technique used for data conditioning for data mining is data warehousing. Data warehousing is used for organizing such unstructured data, it's the middleware that transfer data from the transactional database into a structured, aggregated warehouse. Data warehousing is tasked with data extractions transformations, and load, this is known as the ETL process were the data is modified to be stored in the warehouse. Data in the data warehouse is not like its previous form were it was in the original database, it's an aggregated more cleaned version.

9.3 Human like Intelligence

The goal of today's data mining tools is to reach human experts level, in terms of accuracy and innovation. The promise of such intelligence lies in incorporating more AI techniques into data mining tools. This newfound intelligence will help incorporate data mining into fields that was not usual for such mining to occur. Technically the data mining is one of the main uses of AI algorithms commercially available today among other data-mining related fields.

Such intelligence incorporation has led to fraud detection mining tools, summarization, predictive analysis, and information retrieval tasks to name a few. IBM's SPSS, statistical modeling software, usages many AI techniques, incorporating machine learning also. Data mining seems to be the most prominent frontier were AI is currently thriving. In addition, a new technique rising in the field of AI in data mining is soft computing. *Soft*

computing is considered with computing techniques that tolerate and exploit imprecision, uncertainty, approximation and reasoning. This new and promising technique allows for traceability, robustness, and close resemblance, forming the new term of Machine IQ. *Fuzzy logic* also is a contributor to the advancement of newer more intelligent data mining techniques.

9.4 Interactive Presentation

The trend for the Presentation tools is being more and more interactive with the user. This is due to theadvances in User Interfaces (UI) designs, from graphical interfaces, voice recognition, to touch sensitive displays. This trend of Presentation graphs is called *advanced* Presentation as opposedto the olden types of *static* graphs such as pie charts,histograms and scatter plots. While the interactive -advanced- Presentation tools do have limits such as the need of a multimedia medium such a monitor of a computer, laptop, or a tablet, they are still have the edge of being able to show more complex structures through zooming in and out, 3D rotation, and/or changes in datasets by enabling user input. These types of interactive tools can also be embedded intosystems and websites, due to their nature of being targeted toward end users and able to have multiple outputs.

X. CHALLENGES

Data Mining and Data Presentation is usually more effective if the data on which to be mined are conditioned beforehand. Future directions show the usage of Presentation s output as inputs for Data Mining through the tight integration ofimplementing visual and pattern recognition algorithms in Data Mining functions themselves. Selecting a data mining algorithm can also be challenging. The user must select an algorithm that would represent the set of data accurately; a method to evaluate the representation; and a search criterion.

10.1 Challenges in Data Mining

Currently data mining, in the form we know it today, has not really achieved the potential of what was expected, envisioned in the late 1980's or early 1990's. The vision of becoming a mainstream application, it's widely used but to a degree still limited, data mining hasn't reached that vision. Challenges come in many forms, mainly in three categories, technical, legal, and ethical challenges, all of which they hinder adaption of data mining as a common practice. We shall examine some of these challenges that hinder the further development of data mining.

XI. CONCLUSION

Data mining is a vast, yet an emerging, computer science field. Widely vast and encompassing many other subfields such as *web mining* and *text mining*, and overlaps with fields such as such as *text mining,machine* learning, fuzzy logic, probabilistic reasoning, and computational intelligence. Data mining and Presentation have developed a lot sinceits inception from hundreds of years ago. Many application of data mining have gotten a huge adaption and user base. Google, the internet giant is one of the main adaptors of data mining. Data Presentation today is helping to solve many engineering and scientific problems in ways that were unimaginable before, such as Map-Reducealgorithms.

Future work includes more investigation on the current challenges facing the development and widespread use of data mining and Presentation techniques. The current emerging automated data conditioning tools that provide a more effective dataset to be processed by the data mining tools had an enormous impact on how data mining and Presentation tools are designed. With those emerging tools in mind, data mining and Presentation tools can achieve more than accurate results than before. Also more work should be done in terms of the current ethical issues associated with data mining and Presentation techniques, namely the anonymization problem associated with the privacy concerns of the general public. How to ethically sift through data records without harming or breaching others privacy.

Today, most leading enterprises and organizations depend heavily on such tools for decision support, and business intelligence. Finally, it is clear now how data mining and Presentation tools are essential in the knowledge discovery process, and they have an enormous impact on businesses and the research facilities. While both are separate and have their respectful principals and methods, their integration is eminent for the benefit of both fields.

References

[1] D. Alexander "Data Mining" Internet: http://www.laits.utexas.edu/~norman/BUS. FOR/course.mat/Alex/, [Mar. 11, 2012].

[2] B. Palace, "Data Mining," Internet: http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm, spring 1996 [Feb. 25, 2012].

[3] M. Sousa, M. Mattoso, and N. Ebecken "Data mining: a database perspective," In Proc. *International Conference on Data Mining*, 1998, pp.413-431.

[4] G. Dennis Jr, B. Sherman, D. Hosack, J. Yang, W. Gao, H. Lane, and R. Lempicki "DAVID: Database for Annotation, Presentation , and Integrated Discovery," *Genome Biology*, vol. 4, pp.3-14, August 2003.

[5] V. Friedman, "Data Presentation : ModernApproaches," Internet: http://www.smashingmagazine.com/2007/08/02/data-Presentation -modern- approaches, Aug. 2, 2007 [Mar. 12, 2012].

[6] R. Mikut, and M. Reischl "Data mining tools" *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*,vol. 1, pp.431-443 , September/October 2011.

[7] D. Tegarden, "Business Information Presentation ," *Communications of AIS*, vol.1, January 1999.
[8] S. Few, "Human Perception," Internet: http://www.interaction design.org/encyclopedia/data_Presentation _for_human_perception.html, Sept. 16, 2010 [Mar. 16, 2012].
[9] F. Post, G. Nielson, and G. Bonneau "DataPresentation : the State of the Art," United States of America: Springer, 2002, pp.464.
[10] G. Grinstein, and B. Thuraisingham, "Data Mining and Data Presentation " in Proc. of *the IEEE Presentation '95 Workshop on Database Issues for Data Presentation* ,October 1995, pp.54-56.
[11] D. Keim "Visual Techniques for Exploring Databases," *International Conference on Knowledge Discovery in Databases (KDD '97)*, California, USA, August 1997.
[12] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," in AI Magazine, *American Association for Artificial Intelligence AAAI*, vol. 17, pp. 37-54, Fall 1996.
[13] M. Friendly "A Brief History of Data Presentation ," *Handbook of Computational Statistics: Data Presentation* , vol.2, pp. 15-56, 2008.
[14] E. Tufte, "The Visual Display of Quantitative Information," *Cheshire, CT: Graphics Press*, 1986, pp.200.
[15] S. Allen, "The Value of Many Eyes," Internet: www.interactiondesign.sva.edu/classes/dat aPresentation /updates, Jul. 29, 2010 [Apr.1, 2012].
[16] Kochevar, "Database Management for Data Presentation ," *Database Issues for Data Presentation* , vol.871, pp.107-117, 1994.
[17] V. Friedman, "Data Presentation and Infographics,"Internet: http://www.smashingmagazine.com/2008/ 01/14/monday-inspiration-data- Presentation -and-infographics, Jan. 14,2008 [Jan. 14, 2008].
[18] B. Alpern, and L. Carter "Hyperbox," in Proc. of IEEE Conference on Presentation '91, October 1991, pp. 133-139.
[19] M. Ferreira de Oliveira, "From visual data exploration to visual data mining: a survey," *IEEE Transactions on Presentation and Computer Graphics*, vol. 9, pp. 378-394, July-September 2003.
[20] C. Romero, and S. Ventura "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol.33 (2007) pp. 135–146, 2007.
[21] L. Kurgan and P. Musilek, "A survey of Knowledge Discovery and Data Mining process models," *The Knowledge Engineering Review*, vol. 21, pp. 1- 24,March 2006.
[22] Q. Zhang and R. Segall, "Web Mining: A Survey of Current Research, Techniques, and Software," *International Journal of Information Technology & Decision Making*, vol.7, pp.683-720, December 2008.
[23] G. Pavlopoulos, A. Wegener, and R. Schneider, "A survey of Presentation tools for biological network analysis," *BioData Mining*, vol.1, pp.12, November 2008.
[24] N. Raghavan, "Data mining in e-commerce: A survey," *SADHANA Academy Proceedings in Engineering Sciences*, vol.30, pp.275-289, April-June 2005.
[25] B. Gaddam, D. Ghosh, N. Ahmed, S. Donepudi, and V. Khadilkar, "Computational Intelligence in Data Mining," Internet: http://www.cs.lamar.edu/faculty/disrael/C OSC5100/ComputationalIntelligenceInDat aMining.pdf, [Apr. 1, 2012].
[26] A. Berson, S. Smith, and K. Thearling, "An Overview of Data Mining Techniques," Excerpted from the book *Building Data Mining Applications for CRM*, McGraw Hill: USA, 1999, pp.488.
[27] D. Keim "Pixel-oriented Presentation Techniques for Exploring Very Large Databases," *Journal of Computational and Graphical Statistics*, vol. 5, pp. 58-77, March 1996.
[28] D. Keim, and H. Kriegel "Presentation Techniques for Mining Large Databases: A Comparison" *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, pp.923-938, December 1996.
[29] D. Asimov, "The Grand Tour: A Tool for Viewing Multidimensional Data," *SIAM Journal of Science & Statistical Computing*, vol. 6, pp. 128-143, 1985.
[30] M. Oliveira, and H. Levkowitz "From Visual Data Exploration to Visual Data Mining: A Survey," *IEEE Transactions on Presentation and Computer Graphics*, vol.9, pp. 378 - 394, July-September 2003.
[31] Information Management, "Charting information management how your business works," Internet: www.informationmanagement.com/media/ui/mk2010.pd f,2010 [Apr. 1, 2012].
[32] S. Casner "A Task-Analytic Approach to the Automated Design of Graphic Presentations," *ACM Transactions on Graphics*, vol.10, pp.111–151, April 1991.
[33] D. Plonka, "FlowScan - Network Traffic Flow Presentation and Reporting Tool," *14th Systems Administration Conference (LISA 2000)*, New Orleans, Louisiana, USA, December 3– 8, 2000, pp. 305-317.
[34] M. Marwah, R. Sharma, R. Shih, C. Patel, V. Bhatia, M. Mekanapurath, R. Velumani, and S. Velayudhan, "Presentation and Knowledge Discovery in Sustainable Data Centers," Compute 2009 ACM Bangalore Chapter Compute, Bangalore, India, January 2009. MAIA Intelligence, "Business Intelligence in Manufacturing", 2009, Internet: www.maia-intelligence.com, 2008 [Apr. 1,2012].
[35] M. Ester, H. Kriegel, and J. Sander "SpatialData Mining: A Database Approach" Advances in Spatial Databases, vol. 1262, pp47-66, 1997.
[36] M. Erwig, and R. Gueting, "Explicit Graphs in a Functional Model for Spatial Databases," *IEEE Transactions on Knowledge and Data Engineering*, vol.6, pp.787-803, October 1994.
[37] K. Zeitouni, "A Survey of Spatial Data Mining Methods Databases and Statistics Point of Views," *Information Resources Management Association International Conference IRMA 2000, Data Warehousing and Mining, Anchorage*, Alaska. pp. 229- 242
[38] R. Geary, "The Contiguity Ratio and Statistical Mapping," *Incorporated Statistician*, vol. 5, pp. 115-145, 1954.
[39] S. Chaudhuri, and V. Narasayya, "New Frontiers in Business Intelligence" *The 37th International Conference on Very Large Data Bases*, Seattle, Washington,

pp.1502-1503.
[40] A. Mocanu, D. Litan, S. Olaru, and A. Munteanu "Information Systems in the Knowledge Based Economy" *WSEAS Transactions on Business and Economics*, vol. 7, pp.11-21, January 2010.
[41] T. Ramakrishnan, M. Jones, and A. Sidorova, "Factors influencing business intelligence (bi) data collection strategies: An empirical investigation," *Decision Support Systems*, vol. 52, pp. 486–496, January 2012.
[42] M. Fitzsimons, T. Khabaza, and C. Shearer, "The Application of Rule Induction and Neural Networks for Television Audience Prediction," *In Proceedings of ESOMAR/EMAC/AFM Symposium on Information Based Decision Making in Marketing*, Paris, November 1993, pp. 69-82.
[43] M. Hearst, "What Is Text Mining?" Internet: http://people.ischool.berkeley.edu/~hearst/t ext-mining.html, Oct. 17, 2003 Oct. 17,2003 [May 2, 2012].
[44] K. Cohen KB, L. Hunter, "Getting Started in Text Mining," *Public Library of Science* PLOS, vol. 4, pp.20-22, January 2008.
[45] F. Facca, and P. Lanzi "Mining interesting knowledge from weblogs: a survey," *Data & Knowledge Engineering*, vol.53, pp. 225–241, 2005.
[46] A. Abraham, "Business Intelligence from Web Usage Mining," *Journal of Information & Knowledge Management*, vol. 2, pp. 375-390, 2003.
[47] IBM, "SPSS", Internet: http://www-01.ibm.com/support/docview.wss?uid=swg 21506855, [Apr. 1, 2012].
[48] IBM, "SurfAid Analytics", Internet: http://surfaid.dfw.ibm.com, [Apr. 1, 2012].
[49] Oracle, "Oracle Data Miner 11g Release 2," Internet: http://www.oracle.com/technetwork/databa se/options/odm/dataminerworkflow- 168677.html, Jan. 2012 [Apr. 1, 2012].
[50] SAS, "SAS Enterprise Miner," Internet: http://www.sas.com/technologies/analytics /datamining/mine, Sept. 2, 2010 [Apr. 15, 2012].
[51] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software: an update," *Special Interest Group on Knowledge Discovery and Data Mining SIGKDD Explorer News*, vol. 11, pp. 10-18,June 2009.
[52] IBM, "IBM Parallel Visualizer," Internet: www.pdc.kth.se/training/Talks/SMP/.../ProgEnvCourse.h tm, Sept. 22, 1998 [Apr. 15, 2012].
[53] Cave5D, "Cave5D Release 2.0," Internet: www.mcs.anl.gov/~mickelso/CAVE2.0.ht ml, Aug. 5, 2011 [Apr. 17, 2012].